

Применение методов оптимизации и машинного обучения для выстраивания образовательного процесса

Рябинин Евгений
riabinin.ev@phystech.edu

Project Proposal

Применение метода анализа иерархий для построения рекомендательной системы по подбору направлений обучения для абитуриентов, поступающих в высшее учебное заведение.

1 Идея

Построение рекомендательной системы на основе метода анализа иерархий поможет предсказать предпочтение и проход абитуриента на конкретное направление конкретного ВУЗа. Данная задача является актуальной для большинства ВУЗов ввиду особенностей правил поступления, меняющихся из года в год, а также популярности тех или иных направлений обучения. Создание системы позволит упростить работу приемной комиссии в рамках более качественного отбора абитуриентов по определенной специальности и предсказание их прохода.

1.1 Problem

В статье [1] описана общие особенности применения рекомендательных систем в ВУЗах Российской Федерации. В основе всех методов построения системы рекомендуется использовать вектор характеристик, в частности используются оценки, результаты ЕГЭ, результаты участия в олимпиадах и интересов абитуриента. Ниже приведена иллюстрация (рис.1) упрощенной схемы вектора вычисления характеристик.



Figure 1: Упрощенная схема вычисления вектора характеристик

В статье [2] представлены подобные работы рекомендательной системы с использованием характеристических векторов. Для прорабатывания системы необходимо учитывать как общие правила поступления во все ВУЗы страны на текущий год обучения (см. [сайт](#)), так и особенности правила приема конкретного ВУЗа, например, вот [правила приема](#) в Московский физико-технический институт. В данной задаче будем опираться на правила приема в 2022 году.

Один из методов построения рекомендательной системы в данной области - метод анализа иерархий, в основе которого лежит структурирование задачи на несколько иерархических уровней. В [1] приведено

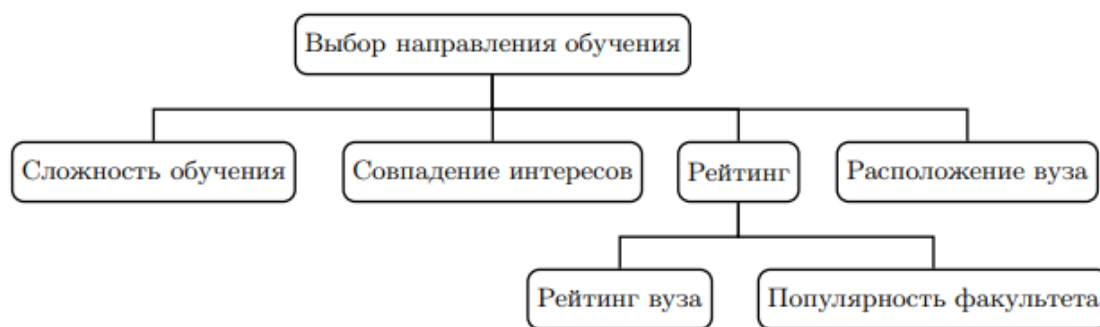


Figure 2: Структура вычисления метода анализа иерархий для выбора направления обучения

краткое описание использования данного метода в образовательных целях. Упрощенная схема представлена на иллюстрации (рис. 2).

Необходимо учитывать наибольшее количество критериев для составления качественной системы конкретного высшего учебного заведения с конкретными направлениями обучения. Одна из задач проекта - составить структуру вычисления метода анализа иерархий с учетом всех предоставляемых данных и правил приема. К сожалению, для составления подобной схемы необходимо учитывать нововведения в системе приемных комиссий и общие тенденции поступления абитуриентов в текущем году. Поэтому схема должна динамически меняться в последующие годы.

Для технической реализации рекомендательной системы потребуется знание машинного обучения и анализа данных [3]. В статье [4] описывается разработка рекомендательной системы для отбора анкет абитуриентов, а также рассматриваются наиболее популярные методы для решения данной проблемы. Авторы статьи рекомендуют в качестве классификаторы использовать "Случайный лес" [5] ввиду высокой точности предсказания.

Вторая задача проекта - выбрать данные по абитуриентам определенного ВУЗа его направления обучения, которые находятся в открытом доступе, представить данные в удобном для обработки виде посредством СУБД, реализовать метод анализа иерархий на языке Python с учетом предоставленной информации и технических возможностей. Списки могут быть найдены посредством [сайта](#).

2 Outcomes

В результате работы над проектом должна быть получена конкретная схема для реализации модели анализа иерархий с обоснованием такого построения. После чего метод должен быть опробован на примере данных по поступлению абитуриентов в 2022 году в одно из высших учебных заведений. На выходе получим ранжированный список абитуриентов по предпочтениям поступления в данное высшее учебное заведение и на данные направления, учитывая общие пожелания, технические данные и вероятность прохода каждого абитуриента. На основе результатов можно сделать вывод о работоспособности или неработоспособности системы относительно конкретных характеристик и описать причину такого поведения.

3 Литературный обзор

В статьях [1] и [6] описывается общие идеи и методы по построению рекомендательной системы для высших учебных заведений по абитуриентам. А в [2] и схожие по тематике представляют решение задачи с подбором рекомендаций с учетом множества различных категорий информации с использованием характеристических векторов.

Про метод анализа иерархий, описанной в статье [1], подробно можно прочитать в следующих книгах: [7] и [8]. Практическое применение данного метода находится, например, в этой статье [9].

Про недостаток данного метода, связанный с возможным нарушением свойства совместности матриц парных сравнений на основе экспертных оценок описано тут [10]. В рамках данной работы применяется автоматический алгоритм заполнения матриц парных сравнений, не нарушающий свойства совместности при использовании корректных данных.

Про применение методов машинного обучения и анализа данных в рекомендательных системах по абитуриентам краткую информацию можно найти в следующей публикации [4]. Там же упомянут наилучший метод для предсказания поступления абитуриентов на конкретную программу, который использует классификатор "Случайный лес". Подробнее про классификатор можно узнать в этой книге [5].

Общая теория по анализу и машинному обучению находится на сайте [3]. В частности, подробности про классификаторы можно прочитать здесь [11]. Также имеются более свежие теоретические и практические материалы по машинному обучению на [данном github](#).

4 Метрики качества

Для оценки качества работы алгоритма мы будем использовать текущие списки поступивших, находящиеся в открытом доступе. Это позволит сравнить работоспособность модели по сравнению с реальным отбором, а именно: совпадение прогнозируемых поступивших абитуриентов с реально поступившими и оценивание вклада конкретных признаков при поступлении абитуриентов в массовом порядке.

При выполнении плана минимума возможно применения более простого метода расстояний [1] или использование иных классификаторов [4], чтобы сравнить с изначальной разработанной рекомендательной системой. Здесь же могут сравниваться подобные характеристики с актуальными данными по поступившим, за счет чего можно будет сделать более точный вывод о корректности работы того или иного реализованного метода.

5 Примерный план

- Изучение метода анализа иерархий: принцип работы, техническая реализация. Изучение материалов по рекомендательным системам в образовательной сфере (приемной кампании абитуриентов), в том числе правила приема в 2022 году. Ознакомление с деталями анализа данных и машинного обучения под данный проект, в том числе вектор характеристик, выбор весов и задачи классификации. Дедлайн: 23 апреля.
- Выбор ВУЗа и направлений (при скором предоставлении доступа буду использовать данные приемной комиссии МФТИ, а направлению обучения рассматривать исключительно по ФПМИ ввиду знакомства с ее внутренней системой, в противном выберу иной ВУЗ, скорее всего ВШЭ), скачивание данных по абитуриентам выбранного ВУЗа за 2022 год. Выбор СУБД для выгрузки данных в удобном формате. Связь СУБД с последующей реализацией алгоритмов на Python. Подготовка Draft 1. Дедлайн: 30 апреля.
- Реализация метода анализа иерархий: создание структуры с учетом специфики данных и техническая реализация. Осуществление связи между данными и алгоритмом, подключение методов машинного обучения. Вывод полученных результатов после применения метода в удобном формате. Подготовка анализа по полученным результатам, основываясь на доступные списки поступивших на конкретные направления данного ВУЗа. Подготовка Draft 2. Дедлайн: 7 мая.
- Завершение работы над проектом. Отлаживание деталей и завершение написания публикационной части. Подготовка к защите проекта. Дедлайн: 12 мая.
- Подготовка выступления к защите проекта. Защита проекта. Дедлайн: 15 мая.
- При выполнении плана минимума и наличии временных ресурсов: изучение метода расстояний. Техническая реализация метода расстояний, привязка к текущим данным и программе. Использование других методов анализа данных и машинного обучения, их реализация. Сравнение характеристик с реализованной программой из плана минимум. Данный пункт может быть реализован до 7 мая.
- Вся проделанная работа в дальнейшем будет использоваться для написания диплома, связанной с темой образовательных процессов.

References

- [1] Артем Валерьевич Белоцкий, Евгений Александрович Суетин. Построение рекомендательной системы по подбору высших учебных заведений для абитуриентов. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления*, (1):66–77, 2016.

- [2] АГ Дьяконов. Алгоритмы для рекомендательной системы: технология lenkor. *Бизнес-информатика*, (1 (19)):32–39, 2012.
- [3] КВ Воронцов. Лекции по машинному обучению. URL: [http://www. machinelearning. ru/\(дата обращения: 16.04. 2023\)](http://www.machinelearning.ru/(дата обращения: 16.04. 2023)), 2023.
- [4] Станислав Сергеевич Рукавишников Валерий Сергеевич Базаркин Евгений Васильевич Малышев, Виктор Владимирович Сливкин. Применение методов машинного обучения для построения рекомендательной системы отбора анкет абитуриентов. *Системы анализа и обработки данных*, (2 (67)):109–119, 2017.
- [5] Leo Breiman. Random forests. *Machine learning*, 45:5–32, 2001.
- [6] Константин Владимирович Митянина, Анастасия Владимировна Гажа. Рекомендательная система "абитуриент" для приемной комиссии вуза. *Проблемы современной науки и образования*, (11 (53)):33–36, 2016.
- [7] T Saati. Prinjatje reshenij. metod analiza ierarhij. М.: *Radio i svjaz*, 278:8, 1993.
- [8] Thomas L Saaty. Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1):83–98, 2008.
- [9] Андрей Анатольевич Байбекова Фатима Наильевна Трунькина Ольга Владимировна Беспалова Наталья Михайловна Сологубова, Лариса Алексеевна Кулаков. Программная реализация принятия решений с помощью метода анализа иерархий. *Вестник Иркутского государственного технического университета*, 22(11 (142)):145–158, 2018.
- [10] В Д Ногин. Упрощенный вариант метода анализа иерархий на основе нелинейной свертки критериев. *Журнал вычислительной математики и математической физики*, 44(7):1261–1270, 2004.
- [11] КВ Воронцов. Метрические алгоритмы классификации. 2013.